AD NUMBER

ADB184716

NEW LIMITATION CHANGE

TO

Approved for public release, distribution unlimited

FROM

Distribution authorized to DoD only; Specific Authority/ Proprietary Info. 7 Jan 94. Other requests shall be referred to Commander, USAMRDAL Command [Provisionl], Attn: SGRD-RMI-, Fort Detrick, Frederick, MD 21702-5012.

AUTHORITY

USAMRMC ltr., 24 Feb 1998

CONTRACT NO: DAMD17-93-C-3150

TITLE: A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING SPEECH INFORMATION FOR SUBSEQUENT PROCESSING BY COMPUTER SYSTEMS

PRINCIPAL INVESTIGATOR: Benjamin Tirabassi, Ph.D.

CONTRACTING ORGANIZATION: Technical Evaluation Research, Inc.
200 White Road, Suite 208
Little Silver, New Jersey 07739-1162

REPORT DATE: January 7, 1994

TYPE OF REPORT: Final Report

PREPARED FOR: U.S. Army Medical Research, Development, Acquisition and Logistics Command (Provisional), Fort Detrick, Frederick, Maryland 21702-5012

PROPRIETARY INFORMATION

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	7 Jan 94	FINAL REPORT (6/15/93 - 1/14/94)

4. TITLE AND SUBTITLE

A Transducer/Equipment System for Capturing Speech Information for Subsequent Processing by Computer Systems

5. FUNDING NUMBERS

C-DAMD17-93-C-3150

6. AUTHOR(S)

Benjamin Tirabassi

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TECHNICAL EVALUATION RESEARCH INC
200 White Road, Suite 208
Little Silver, NJ 07739

8. PERFORMING ORGANIZATION REPORT NUMBER

TR-3150-178

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research, Development, Acquisition and Logistics Command (Provisional), ATTN: SGRD-RMI-S
Fort Detrick, Frederick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

SBIR Phase I

PROPRIETARY INFORMATION

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Distribution authorized to DOD Components only, Specific Authority, January 7, 1994. Other requests shall be referred to the Commander, USAMRDAL Command (Provisional), Fort Detrick, Frederick, MD 21702-5012

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The objective of the Phase I Small Business Innovative Research (SBIR) Program Topic A93-033 is exploratory research to parameterize the speech signal, provide benchmark measurement techniques, and to assemble a superior speech capture system while minimizing the effects of noise and interference. Human verbal response under controlled conditions and when exposed to stress exhibit quantifiable differences that when accessed by computer have direct medical monitoring application. Soldiers answer queries about symptoms, well-being, or perceived capabilities while they are performing their duties. Measures of operational performance for military tasks involving communications such as requestes for fire support or sending of planned target lists are collected by computer systems. These measures demonstrate the feasibility of this transducer/equipment system as a necessary and critical "first step" towards the ultimate application of computer systems to collect and analyze symptoms, moods, and performance data from soldiers in real-time in training centers, the laboratory, and the field with minimal interference to ongoing soldier activities. The algorithms developed for accurate speech capture have demonstrated immunity to a broad range of acoustic and transmission media conditions which include concentrations of acoustic noise in all anticipated regions of the spectrum significant in the human voice response band.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Medical Monitoring, Speech Recognition, Noise Rejection, Speech Capture, Articulation Index, Human Appliance Device	33
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASS	UNCLASS	UNCLASS	Limited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

( ) Where copyrighted material is quoted, permission has been obtained to use such material.

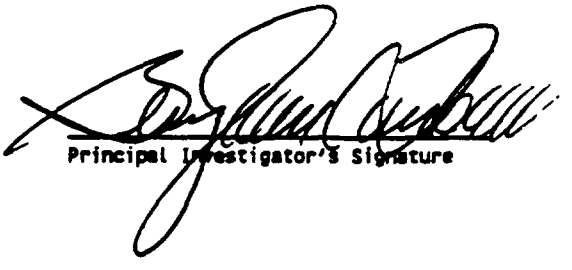
( ) Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

( ) Citations of commercial organizations and trade names in this report do not constitute an official Department of the Army endorsement or approval of the products or services of these organizations.

( ) In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

( ) For the protection of human subjects, the investigator(s) have adhered to policies of applicable Federal Law 32 CFR 219 and 45 CFR 46.

( ) In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.



Principal Investigator's Signature

7 Jan 94

Date

Accesion For	
NTIS CRA&I	☐
DTIC TAB	☒
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
E-4	

A TRANSDUCER/EQUIPMENT SYSTEM FOR CAPTURING SPEECH INFORMATION FOR SUBSEQUENT PROCESSING BY COMPUTER SYSTEMS SBIR PHASE I - FINAL REPORT

FIGURES

1.0 INTRODUCTION

1.1 Summary of Phase I Technical Objectives Met

The objective of the Phase I Small Business Innovative Research (SBIR) Program Topic A93-033, entitled "A Transducer/Equipment System for Capturing Speech Information for Subsequent Processing by Computer Systems" is exploratory research to parameterize the speech signal, provide benchmark measurement techniques, and to assemble a superior speech capture system while minimizing the effects of noise and interference.

Human verbal response under controlled conditions and when exposed to stress exhibit quantifiable differences that when assessed by computer have direct medical monitoring application. Soldiers answer queries about symptoms, well-being, or perceived capabilities while they are performing their duties. Measures of operational performance for military tasks involving communications such as requests for fire support or sending of planned target lists are collected by computer systems. Thus, demonstrating the feasibility of this transducer/equipment system as a necessary and critical "first step" towards the ultimate application of computer systems to collect and analyze symptoms, moods, and performance data from soldiers in real-time in training centers, the laboratory, and the field with minimal interference to ongoing soldier activities. The algorithms developed for accurate speech capture have demonstrated immunity to a broad range of acoustic and transmission media conditions which include concentrations of acoustic noise in all anticipated regions of the spectrum significant in the human voice response band.

Of significance from this research was the development of a benchmark performance methodology and noise characteristics documented for several anticipated environments using different acoustic noise and channel media conditions. This benchmark methodology and characterized noise was the baseline for the research effort to determine speech capture algorithm effectiveness in the anticipated environments. An important part of the research effort was dedicated to the methodology and a description of the techniques proposed for performance testing utilizing a sample of the anticipated noise environments. A prototype system was simulated and designed including algorithm enhancements, speech capture transducers, and necessary ancillary devices for speech intelligibility enhancement and demonstration.

The Phase I research and Phase II advanced development of the superior speech capture transducer/equipment system is demonstrated yielding high quality speech information capture under the following conditions:

- High level ambient noises, 100 db; High intensity impulse acoustic noises, firing of a tank or howitzer;

- High speech recognition accuracy for commercial applications; automated drive thru fast food ordering, automated bank teller machines, train station or airport or stadium kiosks;

- Sensing speech from soldiers who are physically active and determine physical well-being;

- High ambient environmental temperatures and humidities, 38 degrees centigrade and 90% r.h.;

- Pulsed and continuous e-m noise from computers, radios, radars;

- Sensing speech from soldiers wearing various helmets and clothing systems which restrict options for transducer attachment and placement, MOPP1 versus MOPP4;

- The transducer/equipment system is relatively non-invasive and comfortable for the soldier to wear for 12-15 hours, and not require its own separate source of power.

To meet the stated objectives and to help answer the postulated questions, a logical sequence of tasks were performed during the Phase I work. The technical approach consisted of two parallel activities which individually focused on: 1) the melding of proven and novel ways to isolate the signal from the noise to improve speech information content, and 2) development of performance benchmarks to demonstrate the improvements in speech recognition under noisy acoustic conditions and over channels that filter or otherwise interfere with speech signal robust features. This culminated in the simulation and algorithm testing of a prototype system.

The main objective achieved during this SBIR was the improvement of speech capture for automated processing in a tactical or similar noise environment. Phase I research focused on a technique to quantify speech recognition accuracy using an exploratory automated version of the Articulation Index (AI) method. The TERI proposed technique was accomplished using a digital signal processor and specially developed algorithms to assess speech capture performance in terms of the AI. Performance assessment using this technique are accomplished in real-time so that appropriate remedies can be applied to improve the captured speech intelligibility.

As a parallel effort, controlled tests were performed to gauge the effectiveness of different remedies to improve the speech capture. These remedies were performance benchmarked anticipating the Phase II effort which will automate the application of appropriate remedies based upon the "sensed" performance. The prototype system will measure the speech capture performance, in real-time, and apply the appropriate remedy.

We have conducted experiments and speech intelligibility tests using a **Correlative Analysis Algorithm Techniques (PROPRIETARY)** approach that effectively cancels the noise content in speech signal digitized samples. The application of this algorithm has shown significantly improved speech intelligibility in high noise environments above 100 dBA. These research results are documented here and discussed under each specific Phase I executed Task Event.

Results of the research indicates achievement of superior speech capture and word intelligibility using TERI developed algorithms and methods. These algorithms and methods are used to benchmark recognition performance in the presence of high noise environments. TERI considers these algorithms and methods to be **PROPRIETARY** and potentially patentable. This report contains a description of these **PROPRIETARY** algorithms and methods and is subject to safeguarding and non-disclosure in accordance with the terms of our contract DAMD17-93-C-3150 (Provision I.33, I.34 and I.35; I.73). **PROPRIETARY** material is conspicuously labeled and underlined in the text, however, the context with the remainder of the report findings may also be sensitive.

It was found that some remedies, based upon acoustic and electrical noise conditions on the channel, also required mechanical hardware selection for optimum performance. For example, the use of a specific microphone transducer in a particular noise environment was identified with the real-time electronic remedy in combination. We explored the performance of various remedies for a range of signal to noise ratios (SNR) as well as characterizing these tactical noise environments. It has been documented that both the characteristics of the noise, as well as the amplitude, play a large role in the selection of the optimum speech capture remedies.

Phase I research placed heavy emphasis on signal processing remedies for the transducer/equipment speech capture system. It was correctly postulated that significant improvements in speech capture accuracy are achievable through autocorrelative noise cancelling techniques. This signal processing remedy led to the achievement of superior recognition accuracy performance objective in high ambiant noise conditions. These results were achieved using a commercial inexpensive noise cancelling transducer microphone. It is anticipated that Phase II research that investigates other transducer input appliances (i.e., spoon, focal plane, directional, etc.) will improve upon the impressive signal processing results already documented.

It was considered a high priority to develop the benchmark performance techniques prior to addressing the different transducer combinations. Automated data assessment proved more reliable and consistent than the traditional subjective testing methods for speech recognition accuracy. It has also been shown that superior speech capture can be achieved during physical exertion and under stress conditions using speaker independent signal processing algorithms. These algorithms "ignore" speaker variability which achieves the objective for speech capture while wearing

masks, helmets, clothing or in extreme human tolerance environments Speaker independent algorithms continue to perform well in "muffled" and "constricted" environments that distort normal speech patterns.

Results of the SNR testing with the speech capture system have provided significant insight into the way automated recognition systems behave, as compared to human listeners. For a decreasing SNR (increasing noise) human listeners continue to speculate and guess what words are being spoken. This causes a gradual decline in speech capture accuracy as measured over many trials using the AI method. However, an automated signal processing recognizer performs almost error free up to a given threshold SNR, then abruptly ceases to recognize the speech captured (with some measurable variability). This conclusion was drawn from heuristic test data taken at TERI using the VRS-200 speech recognizer and comparing that with AI values for human response found in the ANSI documents.

1.2 Related Work

The TERI principle investigators and key personnel through current speech research and product design effort participate in state-of-the-art speech signal processing and voice recognition disciplines. TERI through the innovative research of these personnel, has contributed to the current success in three key areas related to this proposed effort: 1) The development and implementation of robust speech capture processing algorithms, 2) Voice recognition in a military noise and remote channel media environment, and 3) Development of benchmark methods to quantify the performance of speech capture and encoding systems in terms of speech intelligibility.

A prime example of TERIs cutting edge technology implementation of the best transducer coupled signal processor independent speaker algorithms currently available is exhibited through TERIs development of the VRS-200 stand-alone voice recognition product series. This current knowledge of the state of all related research as embodied in the microprocessor and self-contained power supply is evident by the Phase I success demonstrating superior speech accuracy and noise immunity. It has been through the careful research over the past decade, coupled with constant improvements and enhancements in algorithm and electronic circuitry, that TERI has arrived at its current ability to demonstrate speech capture and recognition in the required noise environments. It has been through careful examination and knowledge growth that TERI has isolated the significant speech channel attributes in terms of algorithms and circuitry necessary to perform the needed filtering, pre-processing, speech parameter coding, template matching. A mature tool set has been developed for the HW/SW embodiment in tactical and rugged solutions. The previous testing achievement and understanding of what combination of transducers perform best is key to the successful conduct of this investigation. This baseline provides stable algorithmic solutions for the investigation of enhancements and provides a basis for assured implementation and availability for Phase I demonstration and the Phase II prototype development and testing effort.

This combination of benchmark performance testing and system engineering knowledge of state-of-the-art algorithms and hardware functions representative of speech channelized media is key to TERIs low risk approach.

TERIs knowledge and awareness of speech capture performance in a noise environment is evidenced by the detailed data corpus for the various critical parameters and awareness of current status in this very specific area of research. TERI has been able to fathom the human factors and physiological frustrations which are well documented in the literature associated with this problem of accurate and reliable voice recognition in a noise environment. TERI also has the added advantage of working with the military in voice recognition and has tested speech capture solutions in the expected noise environments in communications media, vehicular, aircraft, fire battery positions, and under battlefield conditions. TERIs cooperative work with academia and industry as well as various demonstrations in many military applications has provided access to a large database of representative environmental acoustic and signal/noise conditions. Through this experience, TERI can readily identify representative noise conditions and reproduce these conditions in a laboratory environment for the Phase II research for noise immunity solutions. The follow-on capability will be to expedite plans and provide for demonstrations of the prototype enhanced models during the Phase II effort using the 95 percentile human statistics. It has been determined through this active research and implementation in the TERI facility that each of these immunity approaches offer the opportunity for enhancement of the speech transducer/signal processor performance. It has also been determined that the independent speaker knowledge based algorithms and "a priori" parameterization offer superior noise immunity. This phenomena is because of the tendency of the algorithms to "look for" information content thereby ignoring random input signals. The addition of each of the remedies, either singly or in combination, will prove effective in achieving the Phase II goals based upon TERIs previous work in each of the proposed areas addressed in Section 2 (Task Events).

TERIs work in the development of benchmarking techniques and protocol methods for the testing of a speech transducer/signal processor on a quantitative basis is critical past work which is useful to this proposed effort. The literature and experience repeatedly states the inadequacy of current methods to definitively measure the enhancement or detriment of certain noise immunity features when applied to speech intelligibility. TERIs past work in the development of the benchmarking technique previously discussed and the final documentation of this technique under this effort will provide the basis for this formalized and quantitative assessment. TERI personnel past experience and testing provides a good basis for the benchmark methodology development and provides a means to determine the enhancements to speech capture quality in a noise environment in a quantitative manner. It is the combination of all of this TERI related work and current awareness of the state-of-the-art in these key technology topics that is important to the successful achievement of the research goals.

TERI also has the facilities and equipment in place to develop prototype equipments to continue what progress has been successfully demonstrated in previous militarized and ruggedized voice capture, recognition, and communications systems. This provides the long term assurance that the integration of noise immunity devices and the rapid prototyping of algorithms can be achieved with minimum risk. Coordinating these TERI efforts with Aberdeen Proving Grounds, the Naval Post Graduate School, MIT and the Lincoln Laboratories, ARPA, and the Carnegie Mellon Institute researchers has proven valuable in developing and maintaining this awareness of speech processing and voice recognition reality representing the present state-of-the-art. TERI personnel experience in writing technical reports and research papers is amply demonstrated by previous publications and DoD high technology integration contracts.

TERI has also been successful in the commercial field, having developed and marketed the VRS-200 and VRS-1000 product lines. Our engineers have taken innovative conceptual research and followed through the engineering development and production stages. Internal TERI funding has been and will continue to be available to bring new products to the military and commercial marketplace. Most of our products can be classified as signal processing microcircuits. State-of-the-art componentry designed, which is related to this project, includes digital signal processors, flash memory, lattice gate arrays and fine pitch surface mount technology. Knowledge of these technologies is paramount to the design of the proposed micro-electronic implementation of the speech capture system as a single chip product. TERI related work offers the combined experience of innovative research and manufacturing technology to successfully produce the miniaturized speech capture system.

1.3 Relationship With Future Research and Development

The results of the proposed approach is the quantification of both the anticipated noise and channel environment as well as the methodology and quantifiable means to demonstrate performance in that environment. The project starts with a high quality transducer/signal processor and independent speaker recognition as a baseline capability. In a structured and methodical way each of the noise immunity schemes including new algorithms and HW/SW techniques are brought to bear and documented to determine their usefulness and capability to enhance the speech capture quality in a noise environment. TERI has qualified speech capture and computer recognition system products and we have experience in each of the noise immunity techniques. This will provide the opportunity to rapidly approach an optimum combination and proceed to and demonstrate the results of this effort in the minimum amount of time. Software embedded algorithms can be rapidly prototyped and downloaded into the dynamic memory of the VRS-200 product for these trials. The TERI facilities provide a complete and comprehensive resource to integrate the various noise immunity parameterization algorithms and devices in a carefully controlled scientific environment.

The Phase I effort has provided the foundation that has quantified the feasibility of the prototype methods. Now we can proceed with confidence to the Phase II demonstration in the tactical field environment. The results of the Phase I effort will culminated in a description and report on results of rapid prototyping experiments of speech parameterization integrated with the noise immunity software algorithms and ancillary devices. The TERI approach will provide a ready capability to move to the demonstration phase, having earmarked the techniques and methods which offer the most significant enhancements for the tactical noise environment. The proposed benchmark methodology, which quantifies the performance of the VRS in a noise environment, and is a very valuable result of the Phase I effort applicable to all future R&D effort in this field. It is expected that the benchmark methodology will provide a significant advantage for all future research in voice recognition when the results are published and subsequently presented in the various acoustic and voice recognition symposia during Phase II.

Another significant outgrowth of the proposed research is the postulation of a device prototype that will perform in a noise environment as a complete system. Research prior to this proposed effort as documented in the literature has been scattered in pockets of individual research with significant gaps between the research projects associated with speech capture and computer automation using voice commands. The systematic approach proposed here will fill these gaps and bridge the various research projects to provide an integrated solution that is effective in the proposed man-machine interface environment. The proposed controlled and quantified approach will provide all future research in this field with a way to analyze performance and identify error source contributions. The Phase I effort achieved quality voice capture and demonstrated computer recognition with high quality in a noise environment of up to 110 dBA.

The results of this topic research will serve as the building block for future research in the field of speech capture for automated processing. Development of benchmark methods for the assessment of speech capture systems is a significant contribution to this area of research. Using a digital signal processor to automatically quantify speech capture accuracy is innovative to this field which today is prone to subjective and manual assessment methods. The ability of the digital signal processor to accomplish the speech capture assessment, in real time, is significant to future research that will apply remedies to counter noise conditions. The robust nature of the proposed methods will be influential to all speech capture research in either acoustic, thermal or electrical generated noise that would mask the speech signal.

The proposed research will also establish a "high water mark" for the performance of a speech capture system in a high noise environment. This will indicate the current status of the technology in this field of acoustics related to speech automated processing. Publication of these results and benchmark methods would be of immediate and long term benefit to fellow researchers to help standardize the methods and could be a candidate for ANSI review and adoption.

1.4 Potential Post Applications

The application of the proposed project results has potential for use by the military and commercial applications in many areas. Voice recognition has proven to be a successful and efficient method of human-machine interface especially in a complex task and high stress environment. The use of quality speech capture and voice recognition to support the military mission and to improve the performance of the man-in-the-loop has many applications to support crew members in aircraft and in vehicles as well as on board ships, in warehouses, dockside, on the battlefield for the individual soldier, in a High-Mobility Multi-purpose Wheeled Vehicle (HMMWV), and elsewhere. The tactical environment demands quality speech capture and voice recognition in many instances where the hands and eyes of the military crew or soldier may be preoccupied with other tasks thus giving him the additional option to interact with his computerized equipment utilizing voice commands. Typical applications for soldiers include the ability to interface with computers and communications equipment in an environment which requires the wearing of protective gear, gloves, and masks which would preclude use of the traditional I/O devices. Extension of the soldiers or crew member ability to continue to interface with his equipment in darkness as well as in severe weather and noise conditions represents other opportunities for military application to extend his efficiency and ensure successful completion of the assigned missions. The development of enhancements to state-of-the-art voice interaction capability will support the tactical mission in these extended environments and under noise conditions.

A speech capture system that can operate in a noise environment has many commercial applications as well as military. The basic research that will be developed as part of the benchmark capability and the data gathered for each of the HW/SW techniques for computer voice recognition in a noise environment applies to speech intelligibility in general. The art of speech intelligibility is a prime concern in all military and commercial applications be it telephone, radio, communications with computers, or even providing hearing aid enhancements for the hearing impaired. The fundamental research that will be accomplished in speech intelligibility by humans as an interface to machines will help provide a quantifiable way to measure performance and speech intelligibility in a structured and repeatable way. Current literature admits to the gap in speech intelligibility as related to measurable signal to noise ratios. The human ability to "track" speech and discern it in a noise environment will be embodied in the features and techniques which will be investigated in the form of software algorithms and speech segregation transducer electronics. The use of closed end vocabulary sets as well as knowledge based precognitive algorithms will reflect the human ability to anticipate speech input and therefore process and help separate it from the acoustic and channel noise background. The commercial application for the speech capture transducer and signal processor in a noise environment will include direct application to the manufacturing floor, robotics, noisy dock side activity, fire fighting, radio communications, in fact any area where man must communicate with machine in order to support his current efforts.

Speech as a man-machine interface is becoming more and more important in every day activity as the electronics and communications permit miniaturization and portability where none had existed before. The interface between the man and the machine therefore becomes very important to include the use of voice commands in hands and eyes busy activities. Commercial applications also include the military parallel for use over remote channels such as in vehicles and aircraft which are accoustically and electronically noisy environments.

Anticipated use of the basic research also includes the advancement in separation of desirable signals from noise clutter in recorded and transmitted channelized audio information. TERI fully expects that the research and development work carried out here will have application in acoustic exploration and other sensor identification activity as a means to separate needed data from background clutter and other electronic and acoustic signatures. Other applications include voice controlled computers, typewriters, security systems, communications systems, safety devices, inventory controls, robotic control, and a host of other abilities associated with voice command and speaker identification which will be enhanced through this research and development. As technology for voice control systems becomes more advanced such systems will invade nearly every aspect of daily life since voice is the most natural man-machine interface.

Commercial application in the near term have been identified and dialog with several organizations are in progress. Discussions have been held with MIDAS Muffler Co. for speech input of customer service and inventory/warantee into an automated database. Large banking institutions have shown interest in a speech capture system that would operate in a noisy lobby, casino, airport and shopping mall floor for access to the Automated Teller Machines. The provision of a superior speech capture system that has been miniaturized, such as proposed in this advanced development project, has wide commercial application since no prior training is required by the user. The general populace speech variation, accents and anomalies will all be acceptable to the proposed speaker independent speech capture microchip. The small size and relatively inexpensive implementation will make the speech capture system attractive for incorporation in household appliances, tools, and portable convenience accessories that can use voice commands for control or selection. Vending machines, shopping dispenser kiosks, and entertainment virtual reality games of the future will all be voice activated.

2.0 DISCUSSION

2.1 The Work Plan

The work plan is reprinted here as Figure 1, "Schedule of Major Events for Phase I," which identifies the major tasks performed. The initial work of the task provided the formal development of the benchmark performance methodology and a description of the test environment to be followed for the test program. The benchmark methodology, which is described later in this section, was finalized and documented as a quantifiable measure of the

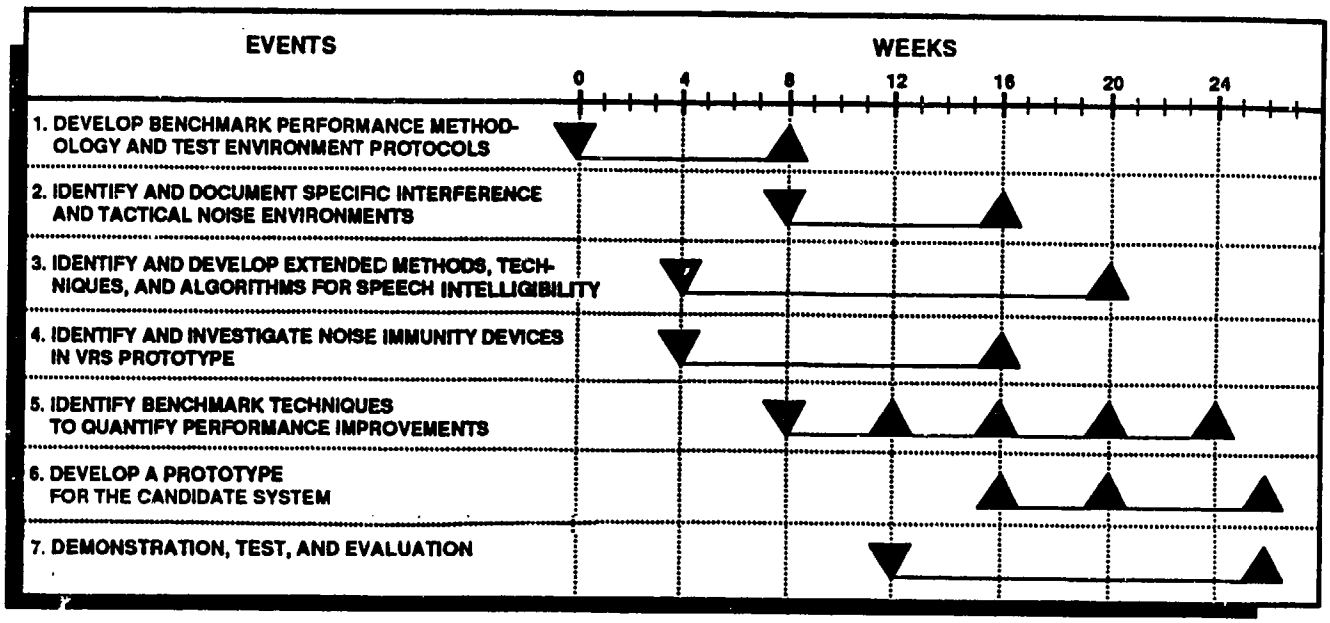


FIGURE 1 SCHEDULE OF MAJOR EVENTS FOR PHASE I

various techniques and algorithm extensions for improved performance in a noise environment. The development of a benchmark technique for the measurement of the improvement of the speech intelligibility system in a noisy environment was significant to the Phase I effort since it represents a serious drawback in the field as documented in the current literature.

2.2 Specific Phase I Results.

Specific research results are provided here by individual Task Event and in a running commentary for each Task Event. These results are readily matched to the Figure 1 schedule of major events for the Phase I research.

Task Event 1: Develop Benchmark Performance Methodology and Test Environment Protocols

During Phase I, effort was expended on a baseline identification and documentation of specific channel, system, transducer interference, and noise environment systems architecture. An important product of this investigative work was the development of the benchmark performance criteria of the enhanced capability in tactical noise environments. It is through the development of this baseline and the quantitative measurement of performance that TERI proposes to leap ahead in performance demonstration and to help solidify an appropriate demonstration methodology and environment for the Phase II effort.

It is also known through research that the nature of the spoken word has a primary effect upon intelligibility. Voiced and unvoiced portions of words, words starting with plosive sounds, and embedded silences; all affect the intelligibility when benchmarked with the AI.

The test protocol requires that a spoken word be presented to the test set-up for measurement of the AI against the established benchmark. Figure 2, "Speech Capture Intelligibility System" presents the test set-up and identifies the functional blocks that affect the capture quality in terms of

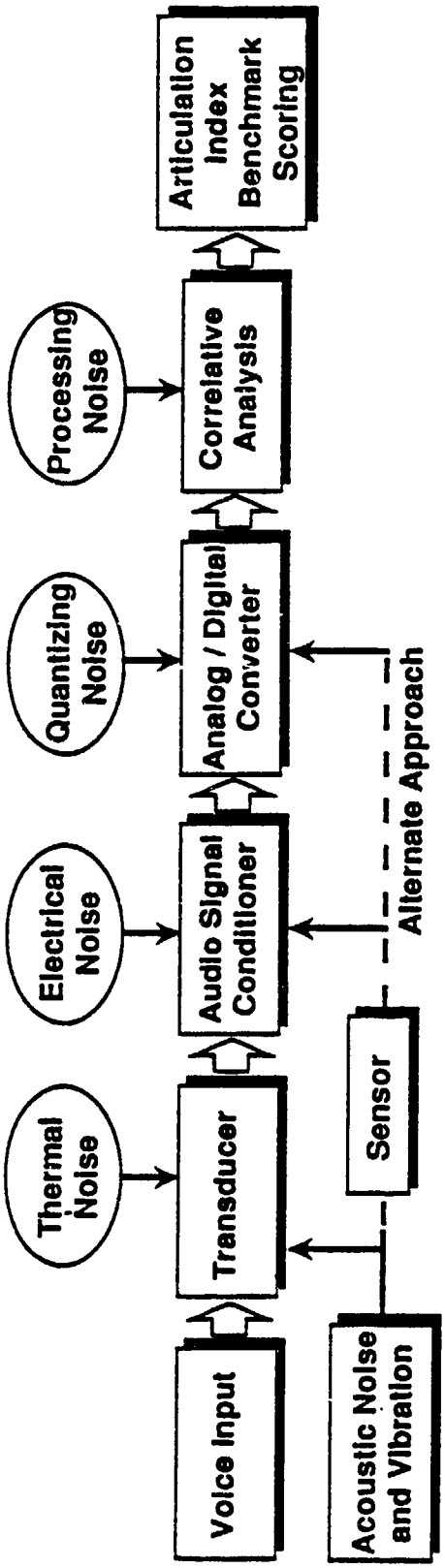


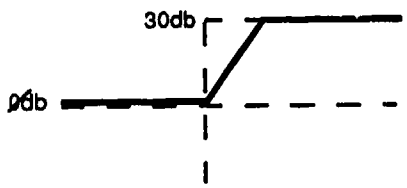
Figure 2. Speech Capture Intelligibility System

intelligibility. Each functional block was implemented using state-of-the-art speech recognition components. This was the baseline experiment which is designed to benchmark a variety of spoken words. The baseline testing included a benign and an increasing amplitude random noise environment. Data was recorded using the AI methodology benchmark tests.

The next set of experiments repeated the baseline tests, however the type of noise was qualified and characterized to represent typical battlefield and tactical environments. These tests were used to examine the robust nature of the signal processing correlation algorithm as a function of different noise characteristic content. The culmination of these initial tests provides the baseline results needed to determine the sensitivity of the postulated correlative assessment approach to AI benchmark. The AI is calculated as follows:

$$AI = \Delta A_{F_1} \cdot C_1 + \Delta A_{F_2} \cdot C_2 + \Delta A_{F_3} \cdot C_3 + \Delta A_{F_4} \cdot C_4 + \Delta A_{F_5} \cdot C_5$$

Where $\Delta A_{F_x} = \sigma (S_{F_x} - N_{F_x})$, (σ) is a non - linear operator as such:



AI CALCULATION FORMULA

The non-linear operational response is such that whenever the difference between the signal to noise is zero or less, assign a value of zero to the difference; and whenever the band pressure level of the signal (speech) exceed the band pressure level of the noise by 30 or more decibels assign a value of 30 to that difference.

The coefficients C1 - C5 are part of the ANSI Standard: they are,

C1 = 0.00024
C2 = 0.0048
C3 = 0.0074
C4 = 0.0109
C5 = 0.0078

PROPRIETARY

The methodology used to benchmark the voice signal capture performance is a Correlative Analysis Algorithm Technique (CAAT). This **PROPRIETARY** CAAT method, as embedded in a digital signal processor (DSP), samples the digitized signal plus noise and uses correlative techniques to separate the signal from

the noise. It is shown that the proposed correlative process effectively separates voice signals from non-stationary (random) noise. Scores are developed by the correlative process that quantify the intelligibility of a particular voice signal. This scoring technique is related to the AI through an heuristic process at various noise levels and gives a benchmark of performance.

An optimized complex 16 point Radix 2 Decimation In Frequency Fast Fourier Transform (DIF FFT) is calculated using-the following steps:

- The power spectrum is RECORDED for each bin of the FFT complex resultants.
- The input data is then windowed with a Hanning window to reduce the leakage error.
- A speech utterance is sampled using a correlating end point detection algorithm.
- The speech utterance is framed at 15.625 Hz periods, across the end points for a variable number of periods. Within each period a 16 point Radix 2 DIF FFT is calculated across the full speech utterance with a RMS value recorded for each octave bin.

The AI is a measure of SNR, however, all systems exhibit internal noise. The system noise will first be determined since the AI is calculated using only the audio signal and noise. The way this is done is to collect an RMS 16 point FFT over a period of time with the audio inputs shunted. It is necessary to also make an audio noise level measurement using the speech channel media (transducer) to characterize the noise (see Task Event 2). This represents an RMS calibration across the response band. These calibration values will be recorded and used in the AI calculation.

Task Event 2: Identify and Document Specific Interference and Tactical Noise Environments

Task 2 was executed to examine the characteristics of various tactical noise environments. A digital tape recording of the Bradley Fighting Vehicle (BFV) noise, recorded at various stations within the vehicle, was obtained from the Human Experimental Laboratory, Aberdeen Proving Grounds, Maryland. This tape was used to continue the investigative analysis of SNR limits of a voice recognizer for these particular noise characteristics. We are pursuing the availability of airborne (helicopter) noise environments to add this data and characteristic analysis to our research effort. A realistic test set-up was used for tactical noise characterization which contained all of the communication channel components shown in the Figure 2 model. Tape recorded BFV (M2) noise was played through the KOSS speaker system as an input to the transducer microphone headset. The transducers served as an input to audio signal conditioner, analog to digital converter and correlative analysis

processing performed by a modified TERI VRS-200 ISA based circuit card. The goal being the characterization of the noise using FFT techniques and cancellation of the noise using autocorrelative statistical processing. Limited testing, which was restricted to BFV noise environmental recordings, showed consistent speech signal extraction at power levels up to 100 dbA (total signal plus noise) as long as the signal exceeded the noise by 15 decibles. The following test equipment and channel media were used to conduct these tests and data recordings.

- Distortion Measurement Set, HP Model 339A
- Multi-Track Recorder-Mixer, Fostex Model 280
- Microphone Transducer, Plantronics Model SDS 1022-03
- Speaker, Koss Model SA30
- Sound Level Meter, RS Model 33-2050.

This effort was truncated and reserved for early Phase II execution in order to concentrate on the signal processing and benchmark research during Phase I.

Task Event 3: Identify and Develop Extended Methods, Techniques, and Algorithms for Speech Intelligibility

Speech parameterization and separation of signal from unwanted superimposed noise can be accomplished in several elements. The algorithms developed for speech parameter separation generally focused on the transducer amplifier, audio signal conditioner, and speech encoder. These algorithms amplifier will include the frequency, time and amplitude domains properly characterized by featurization and then processed to eliminate the unwanted noise and interference. These new concepts in speech parameterization and reconstruction will benefit from this additional dimensionality that current speech capture systems do not use.

Task Event 3 of this SBIR is the development of a "Quantifiable Benchmark for Voice Recognition Systems." The speech accuracy is traditionally calculated using a standard procedure to quantify the performance of the voice recognition system (VRS) based upon previous work at TERI for the US Army Signal School, Ft Gordon ("Concept Evaluation Program #909 Voice Recognition On-The-Move"). A speaker (or number of speakers) is selected and the number of times the speaker utters a response vs the number of misrecognitions is used to determine what the voice recognition system accuracy is in a given environment, (i.e., five misrecognitions for 100 utterances gives a Recognition Accuracy of 95%).

Obviously there is subjectivity, skew, and technique error using this methodology. Also, a large sampling is necessary in order to have a good confidence in the accuracy measurement. This takes a considerable amount of time and patience to ensure controlled conditions. Using this manual

technique, TERI did some test case experiments using at least 25 samples, which gives a 95% confidence (for a absolute score mean of 300 and a absolute score standard deviation of 30) with standard deviation no greater than a 12. Another drawback to the manual technique is that the individual making the utterance may purposely try to make the recognition fail or actually be trained by the machine to pronounce words correctly more often then the casual independent speaker. Lastly the noise conditions may be varying in a biased way and the confidence level of the performance becomes conditional. With this in mind, and the fact that more and more Voice Recognition Systems (VRS) have become available, it is imperative that a simple, quick, and objective method for voice recognition accuracy be developed. The TERI proposed research will automate this with a metric called the AI.

The accuracy measurements are dependant on the particular VRS being tested, whereas the AI is a constant for a particular word (audible signal that is uttered) and is independent of the AI algorithm, (i.e., the standard has already been established (validated)). TERI intends on automating the standard, so that any VRS can be benchmarked using this AI. For example, a particular AI for a system may be 0.6 and have a threshold for that word of 78dBA (random noise crest factor 1.8) and a measured accuracy by mechanically repeating that same word of 100%. This assumes the system has the same strong non-linear decision maker similar to the TERI VRS-200. When the AI is 0.5 for the same word in say a noise of 90dBA the accuracy drops to 0%. Now we know the transition point, and can benchmark the VRS performance for a particular word with repeatable results.

The AI method, as documented in ANSI S3.5-1969 "Methods for Calculation of the Articulation Index," is done manually with analog circuits and bandpass filters. The researched AI method, is calculated using a Fast Fourier Transform (FFT).

PROPRIETARY

The TERI AI benchmarking is done digitally by computer, using TERI developed firmware coding of the FFT and AI transform. This transform is accessed and executed as an Application Program Interface (API) on any suitable DSP.

The AI is calculated under host computer control. The application for the AI flows as such; the user will be prompted to have the background noise source turned on, it will then be sampled under the API. Then the user will be prompted for the utterance to be spoken without the noise source turned on and it will be sampled by the API.

There will be two API types of FFT's (framed and unframed). In each case, eight spectral bins of average values are estimated for use in the FFT. The AI is calculated in two steps, a relatively long FFT spectral estimate will be taken before a word is spoken (unframed FFT API for noise). Then a word is spoken and the spectral estimate for the framed

word is done. Then finally the host computer will calculate the average AI for that given noise environment, by taking the noise signal difference for each FFT bin, multiplying by AI bin weighting coefficient and summing the results.

The embedded algorithm calculates the FFT API's. The sample period is 32 samples at 8 Khz (4 milliseconds) for a 4Khz response (Nyquist). Each of the 32 samples is multiplied by the Hanning window coefficient in order to minimize the leakage error. A Radix 2 Decimation in Time sixteen point FFT will be done. The ANSI AI standard specifies that a Root Mean Square be performed. Because the power spectrum is not constant over the full sampling period it is subject to fluctuations. In order to do the RMS, a series of periodograms will be done, by sliding a "window" for the 32 samples (8 samples at a time (1ms)). An FFT is then performed on each of the 32 windowed samples and the resultant FFT bin samples are squared. The bins are then summed over the full sampling period and divided by the number of windows to get the mean. Finally, the square root is taken of each bin before it is formed as the return message API to the host. The 8 real FFT points are 4khz, 2khz, 500hz, 250hz, 125hz, 62.5hz, 31.25hz, only the first 5 are used for the AI per the standard. More efficient real input FFT's are available using doubling sequences and trigonomic recombination, however, they will be implemented at a later time. The embedded firmware is done using a separate Boot Page of Memory on the VRS-200 DSP prototype.

Task Event 4: Identify and Investigate Noise Immunity Devices in a Voice Recognition System Prototype

Figure 3, "Speech Capture Test Configuration" shows the test set-up for the controlled experiment. Transducers can be benchmarked and substituted at any time. Recording devices are used for test repeatability experimentation to control the signal and noise levels. The recorder is the prime means of characteristic noise playback which will be mixed with the noise signal in various ratios. A signal analyzer is the test tool used to calibrate the input signal and noise content prior to VRS-200 processing.

The test set-up shown has the flexibility to test both live and prerecorded speech input. Separate channels for the speech signal and the noise signal will permit the substitution of alternate transducers, electronic filters, noise canceling algorithms, and other methods as the research activity matures.

Data was recorded for the chosen vocabulary words using characterized noise at different amplitudes. The resultant data was used to determine the robustness of the signal processing algorithm to capture voice input and to separate it from different noise backgrounds.

The first remedy researched for speech autocapture was the use of correlation algorithms to digitally segregate the speech signal from the noise. Using the noise correlative rejection algorithms significantly

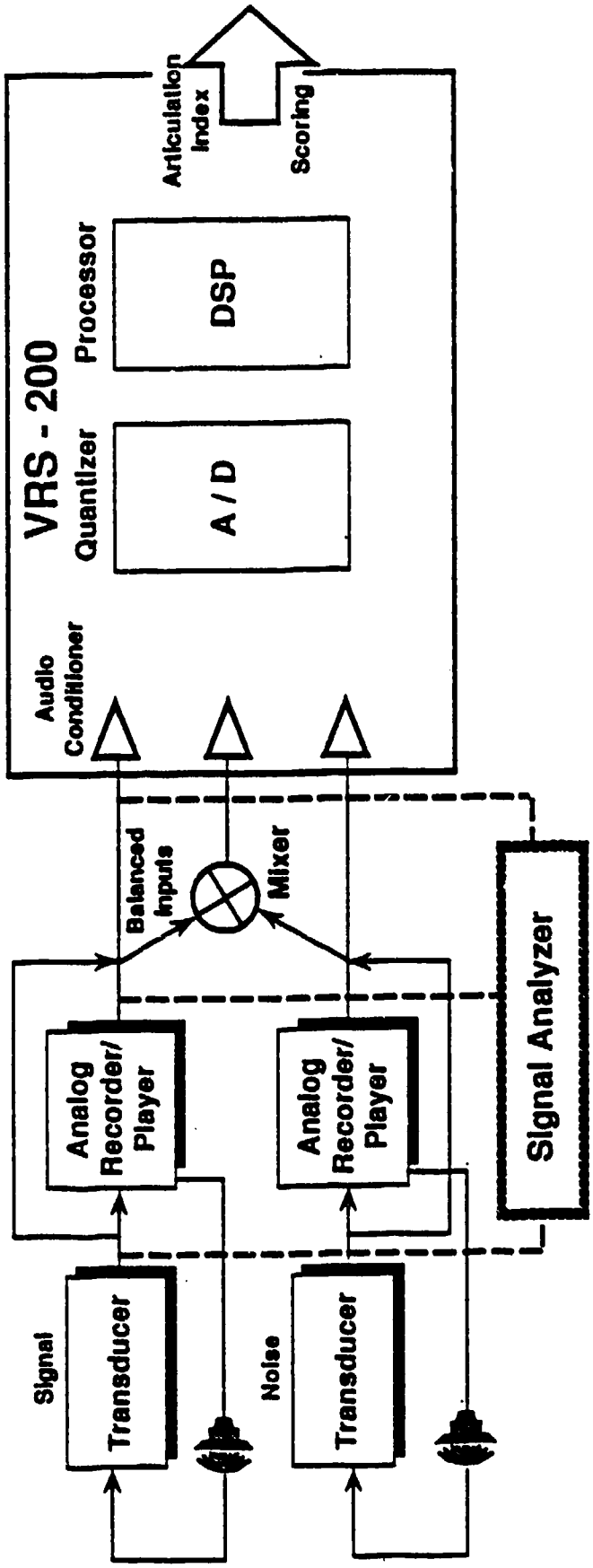


Figure 3. Speech Capture Test Configuration

improved the speech recognizer performance. Very high accuracy scores were recorded in ambient noise levels over 100 dBA. The speech recognizer, using the noise correlation rejection algorithm, still exhibited the rapid decay in speech capture accuracy at critical SNR levels. These critical levels, however, were adjustable by setting the known ambient audio threshold to match the anticipated noise. The ambient threshold is the value of signal plus noise, above which it is assumed there is meaningful speech.

Data was collected by electrically and automatically repeating a selected set of words, this eliminates the human induced skew and technique error and provides a controlled test bed for a large number of sample collections. The data has shown that once the system is set up correctly, with no added noise, the recognition is 100%. Acoustical Speech Recognition Accuracy Versus Noise testing was done after damping adjustments (adjust the distance from the microphone and speaker) all but eliminated the error induced effects. There was a absolute score mean to absolute score standard deviation ratio (variability) of between 4 to 19%, (typically 10%) but NO misrecognitions. This variation can be directly attributable to the inherent system noise, (electrical and processing channel noise) 10% is 20dB for a dynamic range of 110dBA. Now when noise was added there was a very sharp threshold where accuracy went from 100% to 0%. This means that the Voice Recognition System (VRS-200) discriminator is very non-linear. In Figure 4, "Noise Vrs Recognition Accuracy for Various Threshold Settings," the algorithm parameter settings are 50 for the background noise threshold and 300 for the minimum power average contained in an utterance. These are the typical settings for the VRS-200. As is indicated, a sharp transition occurs between 100% accuracy and 0%. In each of the Figure 4 curves, parameters in the algorithm and audio mixing have been changed to determine how parameter settings for the CAAT improve speech capture intelligibility in high noise environments.

Data has been collected for half a dozen different words. The "words" have similar response accuracies but different transition break points in given noise environments. Highly accurate recognition is demonstrated above 100 dBA as shown by the data depicted in Figure 4.

Task Event 5: Identify Benchmark Techniques to Quantify Performance Improvements

There are two basic areas of improvement available in automatic speech recognition: (1) the inherent algorithm development where most attention has been placed on over the years for VRS, and (2) the man-machine interface and appliances. One immediate improvement to the man-machine interface is to use a highly directional microphone and a digital gain control. More elaborate algorithmic designs were used to determine the type of noise in order to modify the recognition algorithm to improve performance. The TERI algorithm uses correlation techniques to perform its signal processing for speech parameter featurization in a noise environment. The inherent nature of the TERI CAAT is to filter out noise using feedback parameters to adjust sensitivity levels of the recognizer.

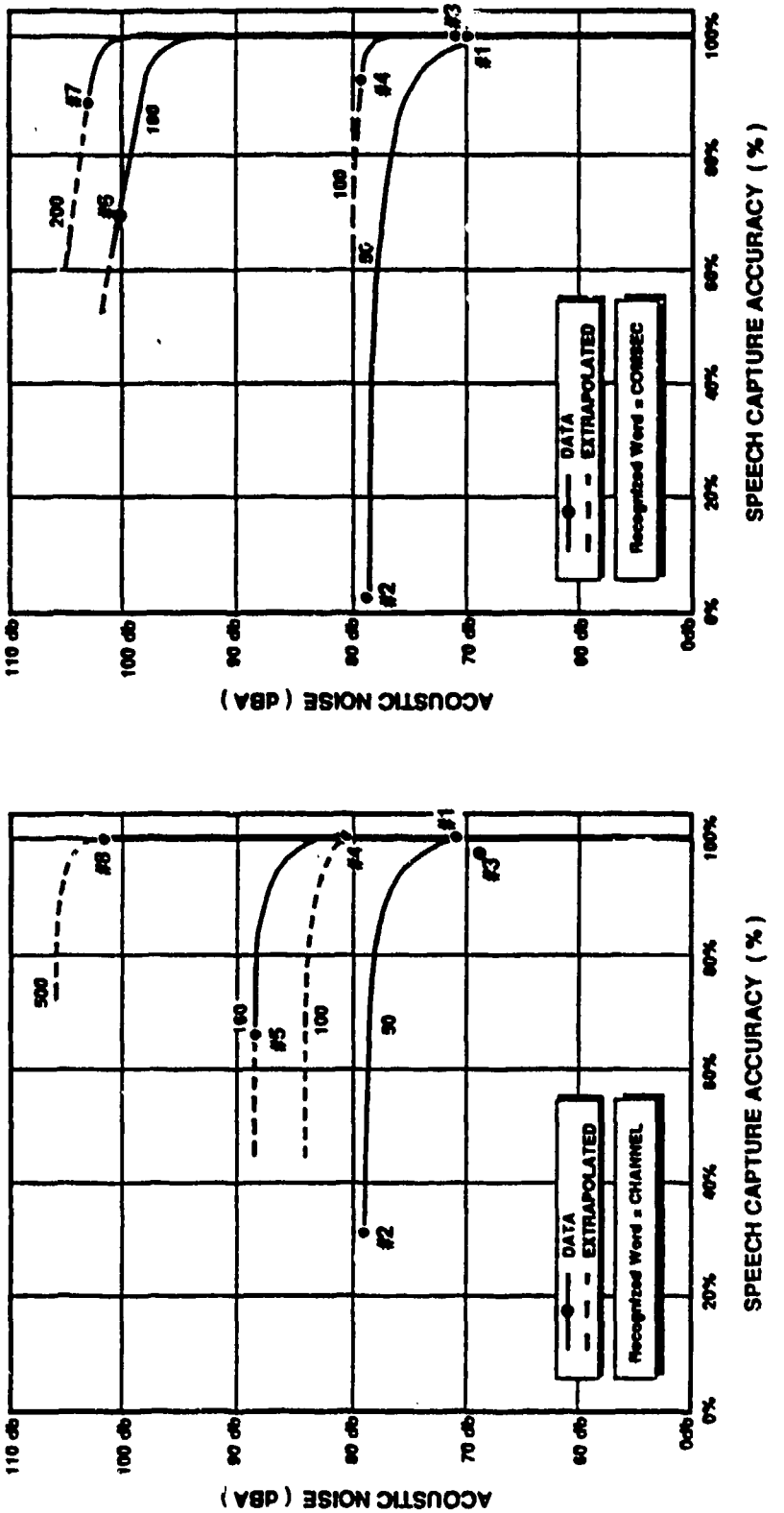


FIGURE 4. NOISE VRS RECOGNITION ACCURACY FOR VARIOUS THRESHOLD SETTINGS (1 of 3)

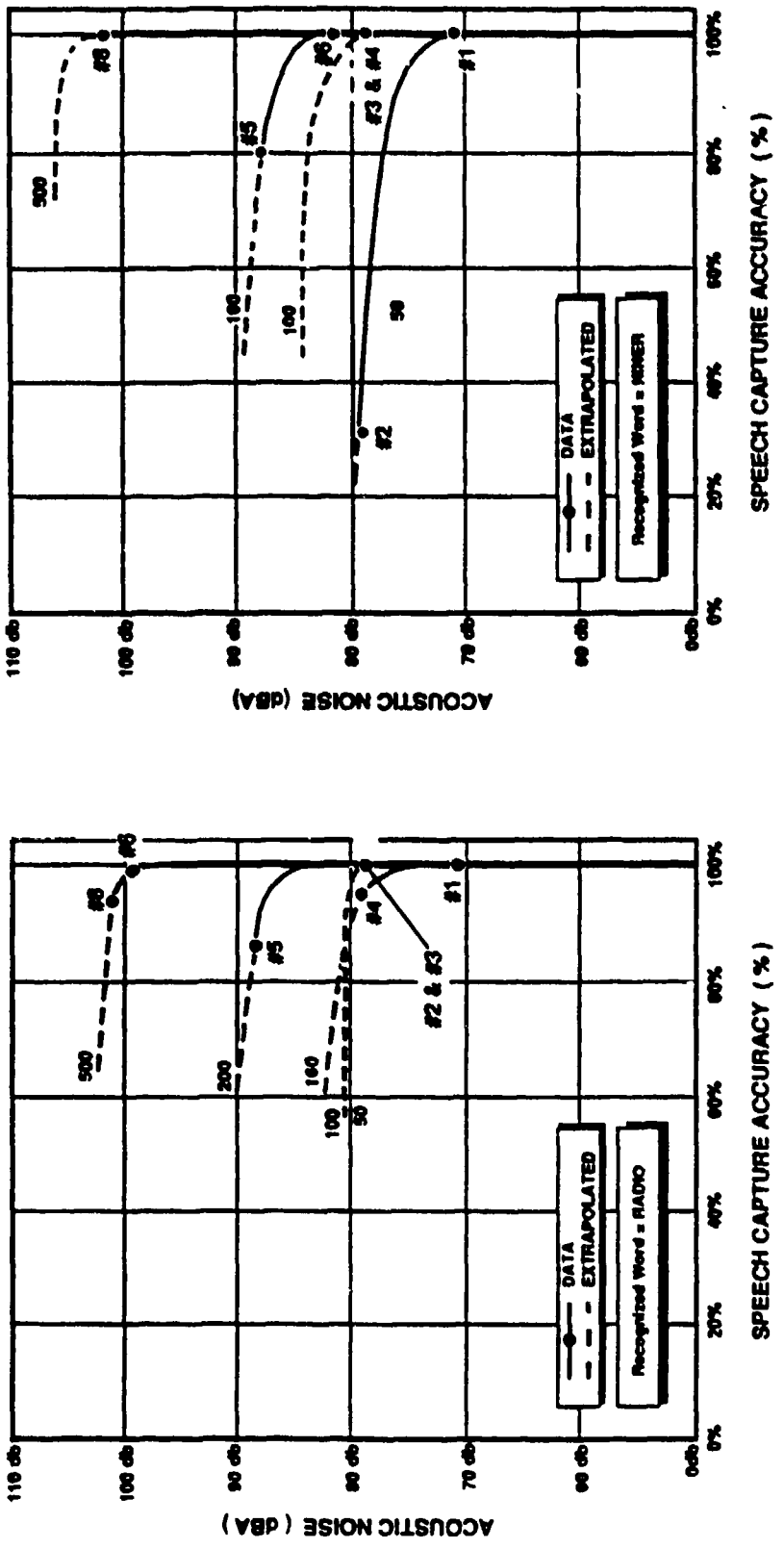


FIGURE 4. NOISE VRS RECOGNITION ACCURACY FOR VARIOUS THRESHOLD SETTINGS (2 of 3)

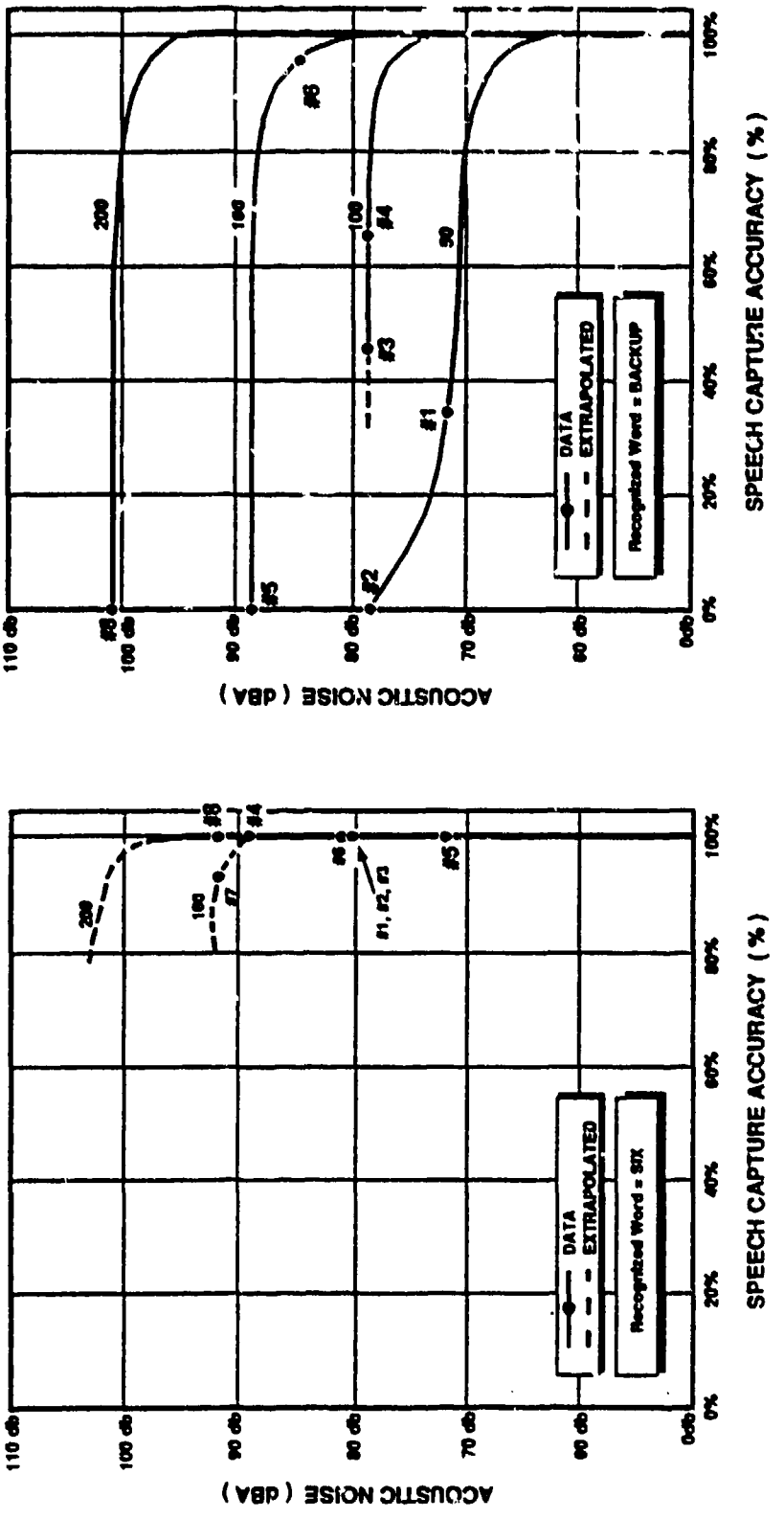


FIGURE 4. NOISE VRS RECOGNITION ACCURACY FOR VARIOUS THRESHOLD SETTINGS (3 of 3)

The autocorrelation algorithm in the TERI CAAT is spectrally estimated over a number of feature periods, each period has a sensitivity to a frequency band to provide the first expectant (autocorrelation) over different sample width periods. Once the stationary noise band(s) are identified then the feature variance bin could be weighted to essentially filter out stationary noise.

Other more sophisticated designs for voice recognition have been investigated, some of them attempt to characterize the essence of the human perception of hearing and have produced unexpected results. Phonetic Fractal Engines, that map the chaotic (large non-linear systems) strange attractors for each phoneme, coupled with back-end Hidden Markov Sets for sequentially (chained) conditional context recognition are future areas for research. Otoacoustic emissions that reflect the human cochlea response to sound produce non-linear signals to the brain. It has been found that the greater the non-linearity of the cochlea response the healthier and more acute the hearing. The degree to which the received audio is distorted provides a measure of the cochlea sensitivity. This is contrary to a generally held premise that purity of the sound is paramount. Human hearing mechanisms produce noise from good pure signals.

This begins to make sense if you consider how the human hearing system evolves: How do we as humans learn to hear? The interesting thing is that if you take a very non-linear transfer function and input a stationary or non-stationary input, the output response is the same! Whereas if the transfer system is linear this is not true. If you consider the "human neural hearing network" as a set of many conditionally weighted nonlinear filters, then we are born with the ability to filter out noise. In other words the neural network is adaptable to an expected noise environment and learns to distinguish this speech signal in expected noise; even though the noise is not present at the time of learning. When the real noise comes along (can't distinguish stationary from non-stationary) the system is already trained to extract the speech signal. The CAAT developed by TERI works in a similar fashion. The speech recognition vocabulary is developed in a benign acoustic environment - then the autocorrelation routines "ignore" the noise in order to extract speech information in noisy environments. This phenomena will be explored further during Phase II in conjunction with the FFT benchmark capability prototype and other speech capture improvements.

Task Event 7: Demonstration, Test and Evaluation

The "a priori" knowledge of noise and its effect on rec ·ition results lends credibility to our postulated solution to meet the o .tive improvements. That is, if we can characterize the unwante oise; then the appropriate remedy or combination of remedies can be automatically applied. In the case of the correlation noise rejection algorithm remedy, it has been shown that significant improvements in speech capture accuracy can be achieved merely by sampling and knowing the dBA level of the noise environment.

The way the speech capture system is envisioned to work is to have the DSP multiplex its function. Alternately, the DSP applies the correlated noise rejection algorithm for speech capture, then perform an FFT to characterize the noise on the "channel". Results of the noise characterization are used to automatically apply appropriate remedies to improve performance. These remedies include adjustment of the ambient level parameters for the correlative noise rejection algorithm based upon the average noise level. Results of the FFT analysis are used to better characterize the noise contributors which can then be used to engage dynamic filters, cancel the noise, or digitally ignore the noise after it is identified.

The FFT real-time execution algorithm is coded and run on the Analog Devices DSP. This algorithm will then be used to automate the benchmarking technique and characterize the various noise samples. It is important to the research that a complete set of data be taken over the range of SNR for a set of chosen words using the speech capture system. This will help calibrate the automated AI calculation compared with the variability demonstrated in results for a set of spoken words.

The FFT algorithm which characterizes the noise and the correlative noise rejection algorithm are the key innovative elements of this research. Integrating these two key elements and implementing them in a single DSP prototype achieves the stated Phase I goals. The remainder of the Phase I effort was dedicated to the implementation of these algorithms and collecting speech capture performance data using the FFT real-time assessment to adjust the correlative noise rejection algorithm.

Effort was also expended during the Phase I effort to document the success achieved in attaining the postulated goals. This record of success is documented in the test data presentation provided as Figure 4 in this report. The Phase II Proposal has been submitted in accordance with the Government furnished instructions and clearly outlines the advanced development of a field test implementation device. The field tests will be structured to gain information and develop confidence in various tactical field noise environments with various transducers. The Phase II effort will incorporate the all important human factor research for suitability, comfort, soldier reaction and formal test and evaluation.

3.0 CONCLUSIONS

3.1 Phase I Results Overview

Technical Evaluation Research Incorporated achieved the Phase I goals by successfully demonstrating accurate speech capture in high ambient noise environments (in excess of 100dBA) using innovative signal processing techniques. This capability was reliable and statistically proven using a new automated and digitally computed bench marking technique that is based upon the Articulation Index method for intelligibility assessment. These

Phase I results have direct applicability to human speech capture and medical monitoring, representing dual use technology advancement in miniaturized signal processing.

Human verbal response under controlled conditions and when exposed to stress exhibit quantifiable differences that when assessed by computer have direct medical monitoring application. Soldiers will answer queries about symptoms, well-being, or perceived capabilities while they are performing their duties. Measures of operational performance for military tasks involving communications such as requests for fire support or sending of planned target lists will be collected by computer systems. Thus, demonstrating the feasibility of this transducer/equipment system as a necessary and critical "first step" towards the ultimate application of computer systems to collect and analyze symptoms, moods, and performance data from soldiers in real-time in training centers, the laboratory, and the field with minimal interference to ongoing soldier activities. The algorithms developed for accurate speech capture have demonstrated immunity to a broad range of acoustic and transmission media conditions which include concentrations of acoustic noise in all anticipated regions of the spectrum which is significant in the human voice response band.

A prototype system was designed to incorporate the superior transducer and signal processing algorithms on a single three (3) inch by five (5) inch printed circuit card using a high speed digital signal processor chip. A sound tape recording at various locations in a moving Bradley Fighting Vehicle was used for the ambient noise during experiments. Other noise characteristics typical in commercial applications such as crowds and loud multiple conversations showed similar signal preservation and noise rejection performance. This innovative research has demonstrated the feasibility of a signal processor based algorithm to assess the ambient noise and then improve the speech signal over a wide range of signal to noise ratio.

The universality of the developed algorithms and signal processing techniques are widely applicable to speech capture intelligibility for subsequent processing by computer systems. Phase I research has culminated in a demonstrable prototype that improves speech capture accuracy, performs speech recognition, and assigns benchmark scores. The benchmark scores are useful in the speech capture performance assessment and to determine the effects of the noise. The prototype design satisfies the goal which prescribes the portable (front-end) capability for the signal processor that is relatively non-invasive and will not require its own separate power source. This evolutionary design will be miniaturized during the Phase II effort and implemented on a single integrated circuit chip. The relatively small chip makes accurate speech capture and subsequent data processing highly attractive for both military and commercial applications. The individual chip solution requires very little power and will be inexpensive to incorporate into human/machine interfaces, speech recognition, and benchmark assessment systems.

3.2 Phase II Technical Objectives

Phase I developed algorithms and methods showed the feasibility of using an automated benchmark technique to quantify performance of a speech capture system. Early Phase II research will apply these methods to other noise environments in order to develop a corpus of data. It has been shown that an understanding of the characteristic noise derived from the automated FFT algorithm is fundamental to eventual speech capture improvement. Specifically, data will be gathered for a variety of input transducers to assess their effect on speech intelligibility. The benchmark algorithms will be used to assess these and other previously proposed remedies. These alternative and combined remedies will then be subject to experiment and data taken to qualify the improvement (or detriment) of these remedies in the presence of various characteristic noise. An important aspect of these experiments will explore the physical practicability and comfort issues addressing physically active soldiers, in high humidity and temperature environments wearing various helmets and MOPP restrictive clothing.

Phase II will make use of the performance benchmark and noise characterization tools developed in Phase I. Use of these tools are essential to the proposed testing of different speech capture appliances and transducers. These AI performance and FFT characterization tools will benchmark the various remedies giving repeatable and reliable data during the proposed field trials. To support these field trials will require incorporation of the automated FFT and performance benchmark algorithms into a portable and self-contained device.

TERI will fabricate several prototype Speech Capture Assessment Devices (SCADs) to be used during the experiments and field trials. These SCADs will gather the benchmark data in highly mobile situations to accurately determine cause and effect in tactical situations. Front-end design of the SCAD will permit rapid modification and exchange of key voice channel devices (like transducer microphones) during the experiment. SCADs will incorporate the automated algorithm "sensing" capability and, in real-time, modify the recognition parameters to optimally adjust to environmental noise characteristics. The combination of hardware and software remedies will provide a comprehensive testbed to further refine the physical suitability of the human interface.

Field trials will be conducted in real tactical environments through coordination with Aberdeen Proving Grounds, MD and currently scheduled Army field exercises. These trials will make maximum use of the SCADs to document the performance and attributes of the various speech capture improvement remedies. Anticipated environments will include: an individual soldier running in MOPP uniform, an computer operator station in a moving tracked vehicle, an operator in a moving wheeled vehicle, a helicopter crew member, and a crew member firing a tank gun or howitzer. This data corpus will be reviewed and documented together with conclusions regarding the achieved speech capture improvement and recommended appliances.

The culmination of Phase II will be marked by the redefinition of a successful speech capture system for automated processing. This refinement will include a further miniaturization design effort in preparation for Phase III. It is important to the commercialization of this product that the implementation be small and inexpensive. This weight and cost factor is also very important to the individual soldier. It is proposed that a singular design that can be universally applied in different tactical situations and noise environments will be a primary goal.

The significance of this Small Business Innovative Research (SBIR) Program Solicitation Topic A93-033, entitled "A Transducer/Equipment System for Capturing Speech Information for Subsequent Processing by Computer Systems" is the exploratory research to parameterize the speech signal, provide measurements, and assemble a noise cancelling or superior speech capture while minimizing the effects of noise and interference. TERIs active participation in state-of-the-art voice processing systems represents a unique opportunity to extend these recognition algorithms which have been demonstrated to be speaker and channel independent with noise immunity. The independent speaker environment also contributes to the successful continuance of speech intelligibility under stress and other media imposed conditions because of the algorithm inherent channel and noise independent qualities. Contributing to the robust nature of TERIs speech processing algorithms is the continued research of the adaptability feature which dynamically changes capture parameters to adapt to acoustic and channel environment. This adaptive feature is selective and permits dynamic algorithm tailoring to support speech accuracy requirements under stress conditions and background acoustic noise. These algorithms have been developed and tailored for specific military use in a broad range of acoustic conditions to include remote communication, armored vehicles, and aircraft. Current speech sensing work has involved coordination with the Army Human Engineering Laboratory (HEL) in Aberdeen, MD and the US Army Communication Electronics Command at Ft Monmouth, NJ for benchmark performance testing with specific application over communications channels and in armored vehicle noise environments.

Performance of speech transducers and signal processing have improved significantly for applications in noise conditions. This basic research will continue the TERI adaptive signal processing and speech parameterization techniques for more robust performance in noise conditions. Additionally the effects of channel media interference will be quantified and superimposed for several expected applications to evaluate the enhanced performance derived from these new concepts.

Soldiers will answer queries about symptoms, well-being, or perceived capabilities while they are performing their duties. Also, measures of operational performance for military tasks involving communications such as requests for fire support or sending of planned target lists will be collected by computer systems. Thus, demonstrating the feasibility of this transducer/equipment system as a necessary and critical "first step" towards the ultimate application of computer systems to collect and analyze symptoms, moods, and performance data from soldiers in real-time in

training centers, the laboratory, and the field with minimal interference to ongoing soldier activities. The algorithms will demonstrate immunity to a broad range of acoustic and transmission media conditions which will include concentrations of acoustic noise in all anticipated regions of the spectrum which is significant in the human voice band.

Of significance from this research will be the development of a benchmark performance methodology and noise conditions to be documented for several anticipated environments using different acoustic noise and channel media conditions. This benchmark methodology and documented noise conditions will be the baseline for the research effort to determine speech capture algorithm effectiveness in the anticipated environments. An important part of the research effort will be dedicated to the methodology and a description of the techniques proposed for performance testing utilizing a sample of the anticipated noise environments. A prototype system will be designed and assembled including algorithm enhancements, transducers, and any necessary ancillary devices for speech intelligibility enhancement and demonstration.

3.3 Phase II Work Plan to Achieve Topic Research Goals

A logical sequence of tasks is proposed for the execution of the Phase II research which continues the successful speech capture experimentation and culminates in a product for commercialization in Phase III. The technical approach consists of: 1) Completing the automated benchmark testing in various noise environments; 2) Investigating the physical appliance portion of the speech capture system; 3) Fabricating self-contained prototypes suitable for field testing; 4) Conducting field trial tests with the prototypes; and 5) Refining the speech capture system design for miniaturization and commercialization.

The work plan is provided as Figure 5, "Overall Phase II Plan - Schedule," which identifies the major tasks to be performed. Sequencing of the proposed tasks takes advantage of the benchmarked performance data in various noise environments to make steady progress in speech capture improvement. We will determine suitable remedies as different transducers and parameter settings are subjected to the controlled experiment. A critical aspect of this experimentation is the "wearability" and "suitability" of the speech appliance to the human in the expected tactical environment. Heavy emphasis will be placed on the human interface experiments during Phase II since the Phase I effort was successful in determining the feasibility of using correlation algorithms to achieve accurate speech capture in 100 dBA noise conditions.

The Phase II development will demonstrate a high quality speech capture system that is well suited to the human interface under the following conditions:

- High level ambient noises, 100db, and high intensity impulse acoustic noises, firing of a tank or howitzer;

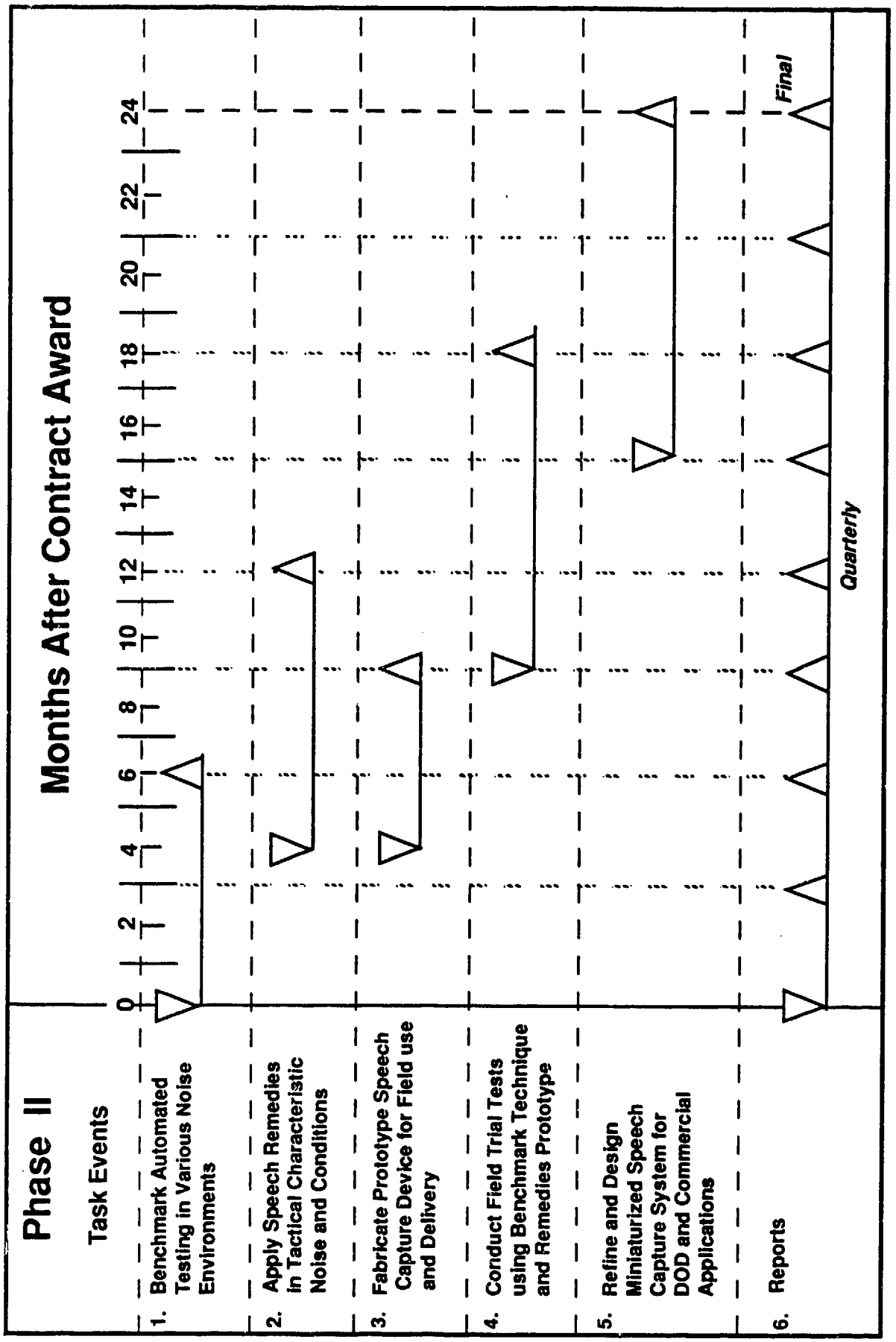


FIGURE 5 OVERALL PHASE II PLAN - SCHEDULE

- Medical monitoring and physical well-being sensor signal processing;
- Sensing speech from soldiers who are physically active; and commercial vending kiosks in open public areas, airports and stadiums;
- High ambient environmental temperatures and humidities, 38 degrees centigrade and 90% r.h.;
- Pulsed and continuous electronic and magnetic noise from computers, radios, radars;
- Sensing speech from soldiers wearing various helmets and clothing systems which restrict options for transducer attachment and placement, MOPP1 versus MOPP4;
- The transducer/equipment system must be relatively non-invasive and comfortable for the soldier to wear for 12-15 hours;
- The system will not use electrical power from an external source;
- Target and demonstrate a commercial application such as Automatic Bank Teller machines and human medical vital sign monitoring.

The following description of the technical approach is presented by Task Event and is designed to match the research topic goals.

TASK EVENT 1: Benchmark Automated Testing in Various Noise Environments

This initial Phase II effort is a continuation of the Phase I work to characterize the various tactical noise environments using the speech capture prototype. The feasibility of using an automated FFT to characterize noise was demonstrated during Phase I. Also it was shown that given some "a priori" knowledge of the noise conditions, it was possible to achieve high quality speech capture. Ability to improve speech capture is predicated upon the characteristics of the noise as well as many other factors. Using the automated benchmark FFT during testing will quantify the noise characteristics for various tactical environments in a controlled and repeatable way.

Continuing this effort as an initial task in Phase II will give the necessary variability in noise characteristic data to investigate alternate speech intelligibility remedies. Phase I experiments were restricted to Bradley Fighting Vehicle noise characteristics, at different SNR values, to quantify the speech capture accuracy using a noise cancelling microphone and noise correlation cancelling algorithm. It is important to characterize other noise sources in a controlled laboratory environment before experimenting with other remedies and in the presence of other uncertainties in the field environment.

This task effort was started during Phase I and its completion is proposed as an extension for project continuity pending Phase II approval. The Phase I effort concentrated on exploring an innovative approach to improve the speech capture accuracy in a high noise environment. This was achieved using signal processing algorithms which have been solidified. Now we propose laboratory experiments to document speech intelligibility improvement using selected transducer appliances and quantifying performance in expanded noise characterized environments.

TASK EVENT 2: Apply Speech Remedies in Characteristic Noise and Environmental Conditions

This task is dedicated to human factor experimentation to determine the effectiveness of various appliance transducers and environmental suitability. A noise cancelling piezo-electric microphone was used for the Phase I electronic algorithm speech capture investigation. This continued effort will experiment with other devices for the human speech input. Interesting microphone candidates include: a spoon-type device similar to that developed for CHIPs, an ear canal accelerometer, a highly directional pick-up microphone, and a focal-plane transducer array composite. There are quality and subjective pros and cons associated with each type proposed. This research will use the speech capture automated benchmark technique to quantify these transducers performance enhancement in the presence of characterized noise, vibration, motion, helmets, masks, sweating, breathing and under human stress conditions.

Performance of various transducers will be benchmarked using the recognition accuracy technique for equivalent AI. A trade-off of performance with operational suitability and wearer comfort is the most significant aspect of this portion of the research. These human factor conclusions will drive the design of the prototype speech capture devices discussed next in Event 3.

TASK EVENT 3: Fabricate Prototype Speech Capture Assessment Device for Field Use and Delivery

The need for the Speech Capture Assessment Device (SCAD) is two-fold: 1) A self-contained and portable speech capture will be used during the field research to benchmark improved performance using different hardware appliances and software algorithms; and 2) A superior speech capture system will be prototyped that incorporates the benefits of the Phase I and Phase II research for delivery to the Government and miniaturized for Phase III commercial production.

A self contained SCAD will be fabricated that integrates the benchmark algorithms and the speech recognition. The implementation will use a single printed circuit card that measures 3" by 5" and is self-contained in a rugged enclosure of similar size. The SCAD will have input and output ports that are compatible with a variety of transducer appliances and has an RS-232 standard computer serial interface.

The TERI VRS-200 is the basic building block of the SCAD which is a single printed circuit card already demonstrated in Phase I to perform in a high noise environment. This circuit card contains analog to digital converter circuitry, a digital signal processor DSP, and flash memory capable of performing as a standalone SCAD system. The VRS-200 circuit card has computer "downloadable" program and data memory which will be modified to incorporate the FFT and AI benchmark algorithms. Additionally, the SCAD will be capable of "front end" speech capture for subsequent automated data processing. Having this dual use, the SCAD can be used to continue the speech improved research and provide feedback for the selection appliances and algorithms for specific tactical noise and environment conditions.

The prototype SCAD (two will be fabricated) will be developed using the research data gathered during the Phase I and the Phase II Event 1 and Event 2. These researched noise characteristic environments and speech capture remedies will be incorporated as hardware interface and software coded techniques. Each SCAD will be capable of being dynamically modified to experiment with different appliances and electronic conditioning to match the field testing projected for Event 4 trials.

TASK EVENT 4: Conduct Field Trial Tests Using Benchmark Technique and Remedies Prototype

TERI proposes to continue the research in the field using the Army Battle Labs to provide user evaluation and suitability data. Use of the SCAD Prototype, developed under this SBIR topic Event 3, is key to the controlled experiment intended in the field. SCAD prototype mobility and self-contained features will be used to good advantage when researching individual soldier human factors issues, whether on foot or in a vehicle.

Research issues of interest include the suitability of the speech capture system in a physically active environment when using various protective clothing, masks and helmets. Non-invasive methods and speech capture appliances are preferred for wearer comfort and freedom of movement. The environmental conditions for testing include high ambient temperature and humidity levels that put severe constraints on speech pick-up transducer choices. Our research will seek to achieve the best trade-off of wearer comfort and speech capture intelligibility. A medical monitor sensor (heart rate or blood pressure) will also be used as a signal source for accurate audio capture and subsequent computer processing. The SCAD controlled benchmarking is an essential experimental tool designed to help conduct this analysis minimizing the electronic and acoustic effects that may mask test results.

A series of field tests using noise cancelling piezo-electric, spoon shaped acoustic, ear canal, etc., speech capture appliances will be tested under similar conditions by a sample set of soldiers. This data will be analyzed and evaluated to determine the advantages (disadvantages) of certain transducer types as compared to others. The SCAD will be adjusted to match the transducer electrical impedance and response curves prior to testing. This feature of the SCAD is important in our efforts to examine the optimal performance possible with these transducer inputs. The SCAD also has the

capability to combine hardware and software remedies in the testing environment. Tests will be conducted to examine and benchmark these combined remedies during the field trials.

All of the proposed testing will contribute to the knowledge corpus in the field of automated speech capture and intelligibility. Transducer performance and human suitability under severe environmental conditions will be documented for different noise conditions. This data is very important to the final recommendation regarding configuration of the Government deliverable prototype speech capture system and subsequent design in anticipation of the Phase III commercially viable product.

TASK EVENT 5: Refine and Design Miniaturized Speech Capture System for DoD and Commercial Applications

All research effort will culminate in the design and specification of a superior speech capture system. This system will embody the successful results of this research in a miniaturized and refined design suitable for the soldier. The proposed speech capture system will represent the leading edge technology in voice interaction under battlefield conditions and simultaneously result in a cost-effective commercial product.

The approach involves the miniaturization of the SCAD prototype desirable features in a single chip package. This miniaturization development can be used as a "front end" to standard Army fielded products (i.e., computers, intercomms, radios, etc.) providing hands-free voice (or medical sensor) interaction under tactical conditions. The miniaturization will reduce power requirements to fractions of a watt so that it may be powered directly from the host system and require no separate power source. This will lighten the soldier's load and provide a commercial viable product to "front-end" even palm top computers with their limited battery power resources.

The refined speech capture system will contain the speech processing adaptability algorithms which dynamically changes capture parameters and adapt to the acoustic and signal channel environment. These algorithms are based upon the research and successful remedy experiments conducted during the field trials. Benchmark performance assessment will be automatically performed and speech capture algorithm parameters modified in real time to adapt to the "sensed" acoustic and environmental conditions on hand. The proposed speech capture system will retain the FFT and benchmarking AI capability to assess the present set of conditions and resulting performance. Additionally, the adaptive signal processing and speech parameterization techniques will be incorporated for more robust performance in noise conditions. The algorithms "downloaded" will demonstrate immunity to a broad range of acoustics and transmission media conditions which include concentrations of characterized acoustic noise in spectrum regions that are significant in the human voice band.

TASK EVENT 6: Reports

TERI proposes to submit quarterly reports that provide research status, conclusions, and progress. Gantt Charts will display actual versus planned progress towards the stated goals for the project. The quarterly reports will also contain the coordination correspondence and results of the field trials with Government points of contact.

A final report and two (2) prototype SCADs will be delivered at the conclusion of the Phase II effort.

DEPARTMENT OF THE ARMY
US ARMY MEDICAL RESEARCH AND MATERIEL COMMAND
504 SCOTT STREET
FORT DETRICK, MARYLAND 21702-5012

REPLY TO
ATTENTION OF:

MCMR-RMI-S (70-1y) 24 Feb 98

MEMORANDUM FOR Administrator, Defense Technical Information Center, ATTN: DTIC-OCP, Fort Belvoir, VA 22060-6218

SUBJECT: Request Change in Distribution Statement

1. The U.S. Army Medical Research and Materiel Command has reexamined the need for the limitation assigned to technical reports written for Contract DAMD17-93-C-3150. Request the limited distribution statement for Accession Document Number ADB184716 be changed to "Approved for public release; distribution unlimited." This report should be released to the National Technical Information Service.

2. Point of contact for this request is Ms. Betty Nelson at DSN 343-7328 or email: betty_nelson@ftdetrck-ccmail.army.mil.

FOR THE COMMANDER:

PHYLLIS M. RINEHART
Deputy Chief of Staff for
Information Management

Completed
15 May 2000
R.W.